

(Rascunho de Artigo para à RAP/FGV) O Efeito do Envelhecimento Populacional no Serviço Militar

Lázaro Damasceno

30 de dezembro de 2025

Resumo

Resumo em PT-BR.

Resumo

Resumo em EN.

1 Introdução

2 Referencial Teórico

3 Metodologia

Python foi a linguagem escolhida para o análise de dados. As bibliotecas polars, matplotlib e scikit-learn foram usadas para a análise dos dados inicial, geração das figuras e análise de dados com ML, respectivamente. A versão do python usada foi o 3.13.11.

O ambiente de desenvolvimento foi o Anaconda na sua versão reduzida – o miniconda. Foi criado um ambiente virtual no miniconda chamado etl_eda_ml para que as bibliotecas do python pudessem ser instaladas.

A escolha reflete as etapas usadas fundamentais e imprescindíveis para a análise de dados: ETL, EDA e ML.

Como passo inicial o ETL (Extract-Transform-Load) engloba a extração dos dados de sua fonte, as transformações necessários para esses estejam funcionais e o seu carregamento para a etapa seguinte, a EDA.

Esta significa análise exploratória de dados. Ela é de suma importância, pois permite descobrir detalhes que apenas recorrendo às abordagens univariada, bivariada e multivariada, poderiam ser notadas.

Exemplos de uso da EDA são, por exemplo, os valores mínimo, máximo e mediano de uma variável para a abordagem univariada; o coeficiente de correlação entre duas variáveis ou mesmo como estas se comportam quando uma terceira variável é incluída na análise.

Na terceira e última etapa, o uso de ML – aprendizado de máquina – permite que o analista de dados consiga superar os limites impostos pelos gráficos estáticos que a EDA gera.

Com ML é possível usar técnicas como regressão, agrupamento (clusterização) e classificação, por exemplo. Assim, será usada a abordagem do aprendizado não-supervisionado para a análise dos dados do serviço militar.

Como consequência, só haverá uma escolha para o algoritmo de agrupamento: ou o KMeans ou o DBSCAN. A escolha será feita com base no resultado do gráfico que mostrará se os pontos formam agrupamentos convexos e globulares ou se apresentam densidade variável e formatos arbitrários.

Contextualizando, usa-se o PCA para reduzir arbitrariamente a dimensionalidade do dataframe, ou seja, reduzir-se o número de variáveis para 2 sem nenhuma medição da qualidade da redução. Esta facilita descobrir qual algoritmo deve ser usado.

Destaca-se que a maneira de descobrir se será usado KMeans ou DBSCAN será gerando um diagrama de dispersão com matplotlib. Caso o gráfico, após a aplicação do PCA, revele clusters bem definidos, arredondados e com densidades similares, o algoritmo KMeans será o mais adequado, dada a sua premissa de minimizar a variância intracluster em torno de um centroide.

Por outro lado, se os pontos assumirem formas alongadas, em arco, ou se houver uma presença significativa de ruído (outliers) que não pertençam a nenhum grupo claro, optar-se-á pelo DBSCAN, que se baseia na densidade para identificar agrupamentos de formatos arbitrários.

No contexto do uso do PCA, a redução da dimensionalidade não é mecânica; ela depende da sensibilidade do analista em interpretar o que os dados tentam comunicar visualmente.

A interpretação dos resultados gerados pelo uso do PCA é de responsabilidade única do analista, pois o algoritmo apenas oferece resultados matematicamente calculados. Assim, interpretar a distribuição dos pontos exige conhecimento técnico necessário para identificar se o KMeans ou o DBSCAN é o mais adequado para os dados do serviço militar.

???

Todas as figuras do Apêndice A tem 15x10 polegadas. No matplotlib, o tamanho da figura envolve a escolha da largura e da altura, ambas em polegadas, respectivamente.

Os dados usados para o desenvolvimento das figuras foi retirado do Instituto

Brasileiro de Geografia e Estatística (IBGE)¹. Este fez a projeção da população brasileira desde o ano 2000 até 2070, com a última atualização da projeção em 2024.

4 Discussão

Esta seção será dividida em duas subseções.

4.1 Dados do Serviço Militar

4.2 Dados da Projeção Populacional

5 Resultados

6 Considerações Finais

¹<https://www.ibge.gov.br/estatisticas/sociais/populacao/9109-projecao-da-populacao.html>

A Figuras da Projeção Populacional do IBGE

Figura 1: Projeção da População por Sexo

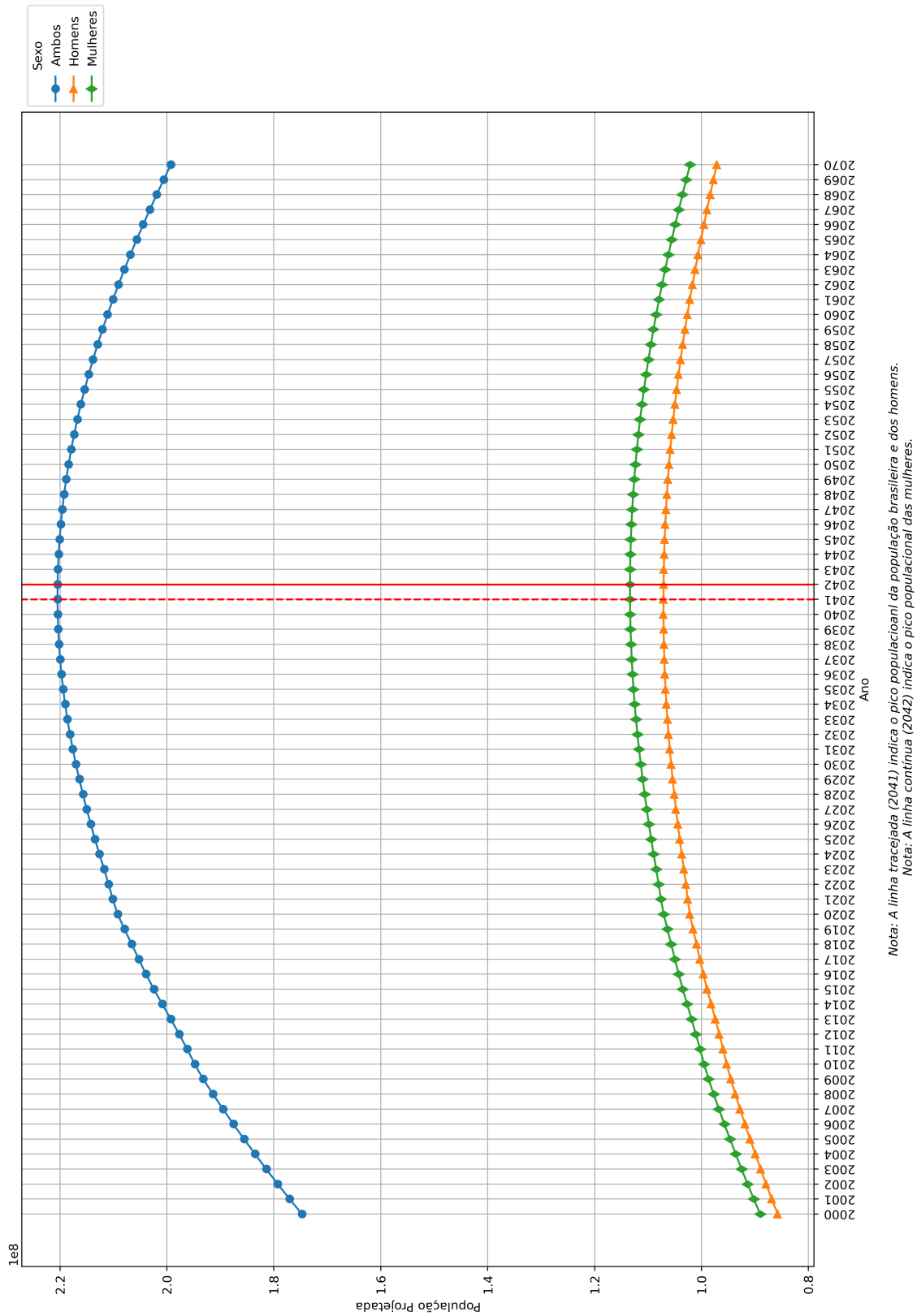


Figura 2: Projeção da População por Grupo Etário

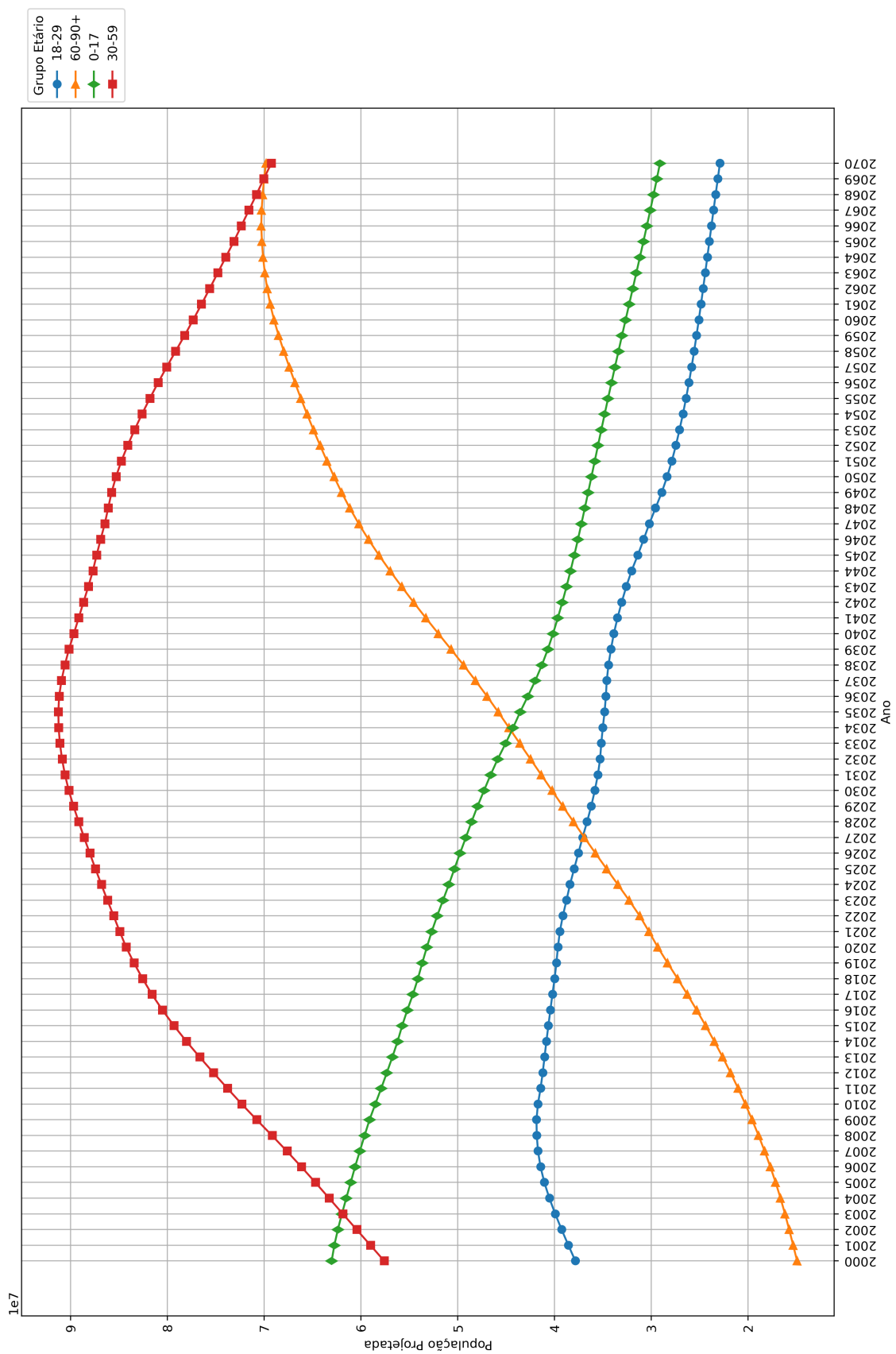


Figura 3: Projeção Populacional Masculina



Figura 4: Projeção Populacional Feminina

